مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

مقارنة اسلوب التحليل التمييزي الضبابي اللبي المكاني مع التحليل التمييزي اللبي المكاني باستعمال المحاكاة

أ.د شروق عبد الرضا سعيد2 م. سكينة شامل جاسم 1
shorouqa.@uokerba.edu.iq 2 Sackineh.sh@uokerbala.edu.iq1
جامعة كربلاء/ كلية الادارة والاقتصاد /قسم الاحصاء1
جامعة كربلاء/ كلية الادارة والاقتصاد /قسم الاحصاء2

المستخلص:-

تعتمد معظم الاساليب الاحصائية ومنها اسلوب التعلم الآلي (Machine Learning) على افتراض ان عينات البيانات المستعملة في التحليل مستقلة ومتماثلة التوزيع والتي يطلق عليها مصطلح ((Identically Independent distributed (iid) لكن هذا الافتراض حول إستقلالية المشاهدات لايتفق مع البيانات المكانية (Spatial Data) وهذه الحالات تحدث في الكثير من المجالات العلمية مثل علم البيئة وتحليل الصور وعلم الاوبئة والدراسات الطبية .. الخ) لانها سوف تفشل في تحديد الارتباط الذاتي المكاني (Spatial Autoregressive).ان اسلوب التحليل التمييزي الضبابي اللبي المكاني (Spatial Kernel Fuzzy Discriminant Analysis (SKFDA)) ياخذ بالاعتبار الاعتمادية المكانية للبيانات من خلال تحديد عدم الدقة في تعيين مصفوفة الانتشار بين الاصناف وداخل الاصناف باستعمال مبدأ المجموعات الضبابية ومقارنة هذا الاسلوب مع اسلوب التحليل التمييزي اللبي المكاني عن طريق مقياسيي الدقة الكلية (Overall Accuracy OA) ومتوسط دقة التصنف (Average Accuracy AA) اذ تم اختيار ثلاثة احجام للعينات داخل كل صنف وهي (50, 200 , 1000) وذلك لمقارنة اداء الاساليب عند حجم العينات الصغيرة والمتوسطة والكبيرة وتحديد عدد الفئات اذ تم تحديد فئتين لغرض تطبيق اساليب التحليل التمييزي باختيار اثنين من النسب لعينات التدريب وهي (0.30, 0.8) و تم اختيار عرض حزمة للمشاهدات ($\mathbf{h_v=700}$) واربع قيم لعرض الحزمة المكاني بشكل عشوائي هي ($\mathbf{h_s=1, 3, 6, 10}$) واختيار دالة لبية كاوسية في هاتيين الاساليب.

الكلمات المفتاحية: الارتباط المكاني ، التحليل الضبابي ،البيانات المكانية، التحليل اللبي ، دقة التصنيف

Comparing the method of spatial kernel fuzzy discriminant analysis with spatial kernel discriminant analysis using simulation

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

Abstract:-

Most statistical methods, including machine learning method, rely on the assumption that the data samples used in the analysis are independent and uniformly distributed, which is called the term (Identically Independent Distributed (iid)). However, this assumption about the independence of observations is not consistent with spatial data. Cases occur in many scientific fields such as ecology, image analysis, epidemiology, medical studies, etc. because they will fail to determine spatial autoregressive correlation. The Spatial Kernel Fuzzy Discriminant Analysis (SKFDA) method It takes into account the spatial reliability of the data by determining the inaccuracy in assigning the diffusion matrix between classes and within classes using the principle of fuzzy sets and comparing this method with the method of spatial core discriminant analysis through two measures of overall accuracy (Overall Accuracy OA) and average classification accuracy (Average Accuracy AA). By choosing three sample sizes within each category, which are (50, 200, 1000), and determining the number of categories, as two categories were identified for the purpose of applying discriminant analysis methods, by choosing two ratios for the training samples, which are (0.30, 0.8), and a bandwidth was chosen for the observations (h_v=700).) and randomly select four spatial bandwidth values (h_s=1, 3, 6, 10) and choose a Gaussian core function in these two methods.

Keywords: Spatial correlation, fuzzy analysis, spatial data, kernel analysis, classification accuracy

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

1- المقدمة (Introduction)

يفترض ان تكون عينة البيانات في معظم خوارزميات لتصنيف التحليل التمييزي اللبي مستقلة وتتوزع بشكل متماثل (i.i.d) مع ذلك غالباً ما يتم عدم تحقق هذا الافتراض في كثير من مشاكل الحياة الواقعية اي بيانات ذات الاحجام الكبيرة (البيانات الحقيقية) اقترح المؤلفون امتداداً لموضوع التحليل التمييزي اللبي Kernel Discriminant Analysis للبيانات المعتمدة مكانياً بناءاً على مقدر الكثافة اللبي (SKDE) والذي ياخذ بنظر الاعتبار المواقع المكانية للبيانات . اذ في عام (2019) اقترح (Boumeddane et al.) خوارزمية تصنيف خاضعة للاشراف جديدة (Supervised Classification Algorithm) تأخذ في الاعتبار الاعتمادية المكانية للبيانات واسموها بمسمى التحليل التمييزي المكاني (Spatial Kernel discriminant Analysis SKDA) والتي تاخذ بنظر الاعتبار الارتباط الذاتي للمشاهدات. و في عام (2020) اقترح (Boumeddane et al.) خوارزمية تصنيف خاضعة للاشراف جديدة (Supervised Classification Algorithm) تأخذ في الاعتبار الاعتمادية المكانية للبيانات والتي اعدوها امتداد للتحليل التمييز اللبي وهي التحليل التمييزي المكاني (Spatial Kernel discriminant وهي خوارزمية خاصة للبيانات التي تعاني من الارتباط الذاتي المكاني بين المشاهدات وتعتمد على التقدير اللبي لدالتي كثافة احتمالية الاولى تتحكم بالقيم المشاهدة , بينما الآخرى تتحكم بالمواقع المكانية للمشاهدات. ولكن في الكثير من الاحيان يكون هذا الاسلوب غير دقيق بسبب عدم الدقة في تحديد مصفوفات الانتشار بين الاصناف وداخل الاصناف لذلك تم استعمال مبدأ المجموعات الضبابية لزيادة الدقة والتصنيف في تحديد هذه المصفوفات، اذ تم استعمال اسلوب تحليل تمييزي ضبابي لبي مكاني (Spatial Kernel Fuzzy Discriminant Analysis (SKFDA)) لزيادة الدقة في النتائج والذي ياخذ بالاعتبار قيم المشاهدات والاعتمادية المكانية للمشاهدات .

2- مشكلة البحث :

تاخذ الاساليب اللبية بنظر الاعتبار عند تقدير دالة الكثافة اللبية اختيار عرض حزمة مناسب لغرض تمهيد البيانات والذي لايكون كافي للتصنيف الدقيق في كثير من الحالات في الواقع الحقيقي والتي في كثير من الاحيان تشوبها عدم الدقة في تحديد قياس وموقع المشاهدة والتي تكون من الأهمية في تحديد مشكلة وتحليل تلك الظاهرة لذلك يجب ان يأخذ بالحسبان الاعتمادية المكانية (Spatial Dependency) للبيانات وذالك من خلال تحليل ياخذ بنظر الاعتبار مواقع المشاهدات

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

والمشاهدات ولابد من تطبيق اساليب تاخذ بنظر الاعتبار قياس المشاهدة وموقعها اضافة الى اساليب اللبية وهذا ما يسمى بالاساليب اللبية المكانية مع الاخذ بنظر الاعتبار الضبابية لزيادة دقة التصنيف وبالتالي الوصول الى اسلوب علمي دقيق وفعال في التصنيف.

3-هدف البحث:

يهدف البحث الى تعامل مع البيانات المكانية التي تعاني من عدم الدقة (الضبابية) لغرض التمييز بين الفئات المكانية المختلفة بدقة عالية و مقارنة اسلوب التحليل التمييزي الضبابي اللبي المكاني مع اسلوب التحليل التمييزي اللبي ومقاييس الدقة (OA, AA) لتوفير تصنيفات دقيقة للعناصر المكانية فضلاً عن قدره هذا الاسلوب في التعامل مع البيانات الضخمة و معالجتها بكفاءة عالية.

4- الارتباط الذاتي المكاني (Spatial Autocorrelation) [6]

يشير الارتباط الذاتي المكاني إلى الدرجة التي ترتبط بها قيم المتغير بقيم نفس المتغير في المواقع القريبة. وبعبارة أخرى، فهو يهتم بدراسة ما إذا كانت القيم المتشابهة تميل إلى التجمع معًا في الفضاء. للارتباط الذاتي المكاني تطبيقات في مجالات مختلفة، بما في ذلك الجغرافيا والبيئة وعلم الجريمة وعلم الأوبئة والاقتصاد. يعد فهم الارتباط الذاتي المكاني أمرًا ضروريًا لاتخاذ قرارات مستنيرة في التحليل المكاني والنمذجة وتخطيط السياسات اذ تلعب العلاقات بين المتغيرات المشار إليها مكانيًا دورًا حاسمًا في التحليل وصنع القرار. ويمكن ان يكون الارتباط الذاتي المكاني موجب ويحدث هذا عندما تتجمع القيم المتشابهة معًا في الفضاء. القيم العالية بحيث يكون قريبة من القيم العالية الأخرى، والقيم المنخفضة قريبة من القيم المنخفضة الأخرى. والارتباط الذاتي المكاني السالب يحدث هذا عندما تتجمع القيم المتباينة معًا في الفضاء. القيم المرتفعة قريبة من القيم المنخفضة، والعكس صحيح. يمكن قياس الارتباط الذاتي المكاني كميًا عن طريق تضمين معلمة الانحدار الذاتي في الانموذج، هنالك مقياس شائع لقياس الارتباط الذاتي المكاني هو احصاءة موران**Moran's Statistics** والتي تكون بالصيغة الآتية:

$$I = \frac{n}{W}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i-\overline{x})(x_j-\overline{x})}{\sum_{i=1}^{n}(x_i-\overline{x})} \quad \dots (1)$$

اذ ان :

I : احصاءة موران

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

n : عدد المشاهدات المكانية

$W = \{w_{ij}\}$: مقياس للأوزان المكانية، وتمثل كمصفوفة أوزان مكانية، والتي تحدد العلاقات المكانية بين المشاهدات.

x_i , x_j : المتغير المدروس في المواقع i وj.

$\overline{x}$: هو متوسط القيم المتغيرة في جميع المواقع.

تتراوح قيم إحصائية Moran's I بين -1 و1 وكما يأتي:

1. اذا كانت قريبة من 1، وهو يشير إلى ارتباط ذاتي مكاني إيجابي، مما يعني أن القيم المتشابهة تميل إلى التجمع معًا. يشير هذا إلى نمط مكاني للتجمع المكاني.
2. اذا كانت قريبة من 1- ، فهي تشير إلى ارتباط ذاتي مكاني سلبي، مما يعني أن القيم المتباينة تميل إلى التجمع معًا. يشير هذا إلى وجود نمط مكاني للتشتت أو التصنيف.

اذا كانت قريبة من 0، مما يشير إلى عدم وجود ارتباط ذاتي مكاني كبير، مما يشير إلى نمط مكاني عشوائي.

5- البيانات المكانية (Spatial Data) [3][2]

تشير البيانات المكانية إلى المعلومات المرتبطة بمواقع جغرافية محددة أو إحداثيات مكانية. تعد البيانات المكانية ضرورية لمختلف المجالات ، بما في ذلك الجغرافيا وعلوم البيئة والتخطيط الحضري والنقل وإدارة الموارد الطبيعية والمزيد.

هناك نوعان أساسيان من البيانات المكانية:

1- بيانات المتجه (Vector Data) : تمثل بيانات المتجه المعالم الجغرافية باستعمال النقاط والخطوط والمضلعات. يتم تحديد كل من هذه الميزات من خلال إحداثياتها المكانية ويمكن أن يكون لها أيضًا سمات مرتبطة بها (مثل الاسم والسكان والارتفاع).

2- البيانات النقطية (Raster Data) : يتم تمثيل البيانات المسحية كشبكة من الخلايا أو وحدات البكسل ، غالبًا ما تُستعمل البيانات النقطية لتمثيل الظواهر المستمرة ، مثل الارتفاع أو درجة الحرارة أو صور القمر الصناعي. لكل بكسل في مجموعة البيانات النقطية قيمة تمثل خاصية معينة في ذلك الموقع.

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

6- التحليل التمييزي (Discriminant Analysis) [4][12]

يعرف التحليل التمييزي ايضاً بإسم تحليل الدالة التمييزية ، هو أسلوب إحصائي يستعمل في مجال التعرف على الأنماط والتعلم الآلي. الهدف الاساسي منه هو تحديد المتغيرات التي تميز بين مجموعتين أو أكثر وتطوير انموذج تنبؤي يعتمد على هذه المتغيرات فقط. تُستخدم هذه التقنية بشكل شائع في مهام التصنيف ، حيث يكون الهدف هو تعيين المشاهدات لفئة أو فئات محددة مسبقًا بناءً على ميزاتها اذ أنه يعتمد على نوع المشكلة ونوع البيانات سواء كانت هذه البيانات كمية او نوعية وله تطبيقات عملية مهمة ، اذ يستعمل في مختلف المجالات من أهمها الزراعية والطبية ، كتصنيف الأمراض ومعرفة شدة الإصابة بها . يعتمد اسلوب التحليل الاحصائي لمتعدد المتغيرات على الظواهر التي لها ابعاد ومتغيرات متعددة لوصفها وتحليلها ويهتم بمسألة التمييز بين مجموعتين او اكثر والتي تشترك فيما بينها بمجموعة من الصفات والخصائص بدرجة مختلفة وذلك باستعمال دالة خاصة تسمى الدالة التمييزية (Discriminant Function). إن عملية التصنيف هي عملية تلي عملية تكوين الدالة التمييزية، إذ يتم الاعتماد على هذه الدالة في التنبوء وتصنيف المفردة الجديدة لإحدى المجموعات قيد الدراسة بأقل خطأ تصنيف ممكن.

7-الدوال التمييزية (Discriminant Functions) [11]

تبنى الدالة التمييزية على فكرة أساسية وهي تقسيم المشاهدات الى مجموعتين وذلك بالاعتماد على مجموعة من المتغيرات أو العوامل إذ تعمل على زيادة درجة التجانس بين مفردات المجموعة الواحدة وتقليل درجة التجانس بين المجموعتين وبالتالي تسهيل إمكانية تصنيف مشاهدة جديدة الى احدى المجموعتين بأقل خطأ تصنيف

* دالة التمييز الخطي (Linear Discriminant Function) [11] [5]

الدالة التمييزية الخطية (Linear Discriminant Function (LDF)) هي إنموذج رياضي يستعمل في التحليل التمييزي الخطي (LDA) للتمييز بين الفئات المتعددة من خلال إيجاد مجموعة خطية من المييزات التي تفصل بين الفئات إلى أقصى حد . تهدف الدالة التمييزية الخطية إلى إيجاد متجه مميز مثل (W) يزيد من نسبة التشتت بين الاصناف ويقلل الانتشار داخل الصنف الواحد وهذا يسمح للفصل الفعال للفئة في مساحة ذات أبعاد أقل ومن المهم ملاحظة أن (LDA) تفترض أن الفئات لها مصفوفات تغاير متساوية تقريبًا وأن المييزات (المتغيرات داخل كل صنف) تتوزع توزيعاً طبيعياً.

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

ليكن لدينا المتجه X بالبعد d الذي يمثل قاعدة بيانات لمجموعة من الميزات (Features) فانه لحساب الدالة التمييزية الخطية نتبع الخطوات الآتية:

1- نحسب متوسط كل صنف (μ_i) ومصفوفة التباين المشترك (Σ) لكل صنف.

2- نحسب مصفوفة الانتشار المجمعة داخل الاصناف (S_i) عن طريق جمع مصفوفات التباين المشترك لكل صنف على حدة.

$$S_i = \sum_{j=1}^{v} \Sigma \quad (for\ each\ class\ i) \quad \dots (2)$$

اذ ان:-

Σ : تمثل مصفوفة التباين المشترك لكل صنف i

3- نحسب المتوسط العام (μ) كمتوسط لجميع متوسطات الاصناف لكل صنف.

4- نحسب مصفوفة الانتشار بين الاصناف S_B باستعمال الفروق بين المتوسطات وكالآتي:

$$S_B = \sum_{i=1}^{v} N_i \left(\mu_i - \mu\right)\left(\mu_i - \mu\right)' \quad (for\ each\ class\ i) \quad \dots (3)$$

اذ ان :

N_i : عدد العينات في الصنف i

μ : المتوسط العام لجميع الاصناف

μ_i : متوسط كل صنف i

5- نحل المعادلة الاتية :

$$S^*{}_B W = \lambda S_i \quad \dots (4)$$

اذ ان :

W : المتجه المميز (Eigen Vector)

λ : الجذر المميز (Eigen Root)

ويستعمل المتجه المميز (W) لاسقاط نقاط بيانات جديدة على الخط المستوي لغرض تصنيف المشاهدات عن طريق الدالة الآتية :

$$LDF = W'X \quad \dots (5)$$

)

ونصف نقاط البيانات الجديدة بناء على قيم (LDF) .

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

8-التحليل التمييزي اللبي المكاني (Spatial Kernel Discriminant Analysis) [9] [7]

يفترض ان تكون عينة البيانات في معظم خوارزميات التصنيف التحليل التمييزي اللبي مستقلة ومتماثلة التوزيع (iid) ولكن لايتحقق هذه الافتراض في كثير من مشاكل الحياة الواقعية.

يعد التحليل التمييزي اللبي المكاني (SKDA) امتداداً لموضوع التحليل التمييزي اللبي (KDA) الذي ياخذ بالاعتبار البيانات المعتمدة مكانياً بناءاً على مقدر الكثافة اللبي (SKDE) والذي اقترحه الباحث (Dabo- Niang et. al, 2014) والذي ياخذ بعين الاعتبار المواقع المكانية للبيانات اذ تسمح خوارزمية التصنيف الخاضعة للاشراف بتصنيف المشاهدات في المواقع الجديدة التي لا تعتمد فقط على المشاهدات نفسها بل على موقع المشاهدة ايضاً، بمعنى اخر يعتمد التصنيف على المشاهدات في المواقع المجاورة باستعمال جميع بيانات التصنيف ، قيمة المشاهدات, وموقع المشاهدة الجديدة ونموذج التدريب من المواقع القريبة .

وبالتالي فاذا كانت مسافة بين موقعين عالية فان قيمتها تكون مستقلة ومن المرجح فانها تنتمي الى الفئات المختلفة فبذلك فان قاعدة التصنيف تتمكن من تخصيص المشاهدات الجديدة $(x \in R^d)$ في المواقع (i_0) والفئة (k_0) .

اذ ان:

$$\dots (k_0 = arg^{max}_{k\in\{1,2,\dots,m\}}(\hat{\pi}_k \hat{f}_k(x)) \quad (6$$

وان $\hat{\pi}_k$ يقدر وفق الصيغة الآتية :

$$\dots (\hat{\pi}_k = \frac{m_k}{\hat{N}} \quad (7$$

لنفترض ان $(X_{k1}, X_{K2}, \dots, X_{kmk})$ هي مشاهدات بالبعد d بـ (k^{th}) فئة . اذ يتم تحديد مقدر الكثافة اللبي المكاني حسب الصيغة الآتية:

$$\dots (\hat{f}_{k(x_j)} = \frac{1}{m_k h_V^d h_s^2} \sum_{i\in I_n}{}_{c(X_{ki})=k} K_1\left(\frac{x_j - X_{ki}}{h_v}\right) K_2\left(\frac{||i-j||}{h_s}\right) \quad (8$$

اذ ان :-

C(.) : هي صنف فئة من المشاهدات .

m_k : هي حجم الفئة k^{th} لبيانات التدريب.

$\hat{N}$: هو حجم مجموعة التدريب .

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

K_1 : هي الدالة اللبية متعدد المتغيرات التي لها قيم في فضاء الاعداد الحقيقية ذو البعد R^d d .وتعرف على انها دالة لبية مضاعفة (Multiplicative Kernel) وكالآتي:

لكل $X = (x_1, x_2, \dots, x_d)$ في R^d فان:

$$K_1(x) = K(x_1) \times K(x_2) \times \dots \times K(x_d) \dots (9$$

اذ ان k الدالة اللبية لاحادي المتغير .

وبصورة عامة فان دالة الكثافة اللبية متعددة المتغيرات يمكن ان تكتب بالشكل الآتي:

$$K_1\left(\frac{x_j - X_i}{h_v}\right) = K\left(\frac{x_{j1} - X_{i1}}{h_v}\right) \times \dots \times K\left(\frac{x_{jd} - X_{id}}{h_v}\right) = \prod_{i=1}^{d} K\left(\frac{x_{j1} - X_{i1}}{h_v}\right) \dots (10$$

9-خوارزمية الجار الأقرب الصنف k الضبابية [1][4]

(Fuzzy k-Nearst Neighbor Algorithem) (FKNN)

تعد هذه الخوارزمية امتدادًا لخوارزمية الجار الأقرب ذات الصنف K k-Nearest Neighbors ((KNN)) التقليدية التي تتضمن مفهوم الضبابية للتعامل بشكل أفضل مع نقاط البيانات غير المؤكدة أو المتداخلة. تقوم خوارزمية ((k-Nearest Neighbors (KNN) بتعيين نقطة بيانات جديدة لفئة الأغلبية بين أقرب جيرانها في الصنف k أي انه إذا كانت هناك نقطة في مساحة، فإن النقطة الأقرب إليها هي النقطة التي تمتلك أقل مسافة منها ، اما خوارزمية (FKNN) تقوم بتعيين قيم الانتماء لفئات متعددة لكل نقطة بيانات مما يشير إلى درجة الانتماء إلى تلك الفئات. اذ يتم في هذا الاسلوب الحصول على درجات الانتماء المجهولة وكل مركز من خلال خوارزمية الاساسية للجار الاقرب للصنف k الضبابية وتكون كالاتي:

1) تحديد قيمة k وهي المصنفة المجاورة التي سيتم استعمالها في التصنيف.
2) حساب مصفوفات قياس مختلفة بين أزواج من متجهات الاصناف في مجموعة التدريب.
3) تعيين العناصر القطرية لمصفوفات القياس المختلفة على أنها ما لا نهاية.
4) نرتب تصاعدياً مصفوفة المسافة بحيث تتعامل مع كل عمود من أعمدتها على حدة ثم تجميع الفئات للأنماط الموجودة المتجاورة.
5) نحسب درجة الانتماء الفئة i للنمط j باستعمال الصيغة الآتية:

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

$$u_{ij} = \begin{cases} 0.51 + 0.49\text{x}\left(\frac{n_{ij}}{k}\right) & if\ i\ is\ the\ same\ as\ j^{th}\ lable\ of\ the\ pattren \\ 0.49\text{x}\left(\frac{n_{ij}}{k}\right) & otherwise \end{cases} \quad \ldots(11)$$

اذ ان :

n_{ij} : يمثل عدد الانماط المتجاورة للنمط j التي تنتمي للفئة i

بحيث ان u_{ij} تحقق الخصائص الآتية:

$$\sum_{i=1}^{c} u_{ij} = 1 \quad -1$$

$$0 < \sum_{i=1}^{c} u_{ij} < \text{N} \quad -2$$

هذا يعني ان مصفوفة الانتماء الضبابية تكون كالآتي:

$$\ldots (12)\ U = [u_{ij}](i = 1,2, \ldots . c, j = 1,2, \ldots, N)$$

10-التحليل التمييزي الضبابي اللبي المكاني

(Spatial Kernel Fuzzy Discriminant Analysis)

يتم تحديد مصفوفة الانتماء الضبابية U بواسطة قياس كل الميزات وكل عينات التدريب اي التحكم في درجة الانتماء الضبابية بدمج درجة الانتماء الضبابية في تعريف مصفوفة الانتشار بين الفئات وداخل الفئات لفضاء الميزة عالي الابعاد Â واضافة المسافة الاقليدية بين الموقعين i و j وبعد ذلك يمكن تنفيذ التحليل التمييزي الخطي (LDA) بتعظيم معيار (Fisher) كآلاتي:

$$\ldots (13)\ \max_{v} J(v) = \frac{v^T S_{SFb}^{\varnothing} v}{v^T S_{SFt}^{\varnothing} v}$$

اذ ان:

$$\ldots (14) S_{SFb}^{\varnothing} = \frac{1}{N}\sum_{i=1}^{c} n_i \left(\tilde{m}_i^{\varnothing} - m_o^{\varnothing}\right)\left(\tilde{m}_i^{\varnothing} - m_o^{\varnothing}\right)' ||i - j||$$

والمعادله اعلاه تمثل مصفوفة الانتشار الضبابية المكانية بين الاصناف في فضاء الميزات Â

$$\ldots (15) S_{SFt}^{\varnothing} = \frac{1}{N}\sum_{j=1}^{N}\left(\Phi(x_j) - m_o^{\varnothing}\right)\left(\Phi(x_j) - m_o^{\varnothing}\right)' ||i - j||$$

اذ تمثل مصفوفة الانتشار الضبابية المكانية الكلية في فضاء الميزات Â

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

اذ ان :

$m_o^{\varnothing}$: متجه المتوسطات لعينات التدريب في الصنف i .

فان الحل الامثل للمعادلة (13) يمكن الحصول عليه من التحويل الى معادلة القيم المميزة الآتية:

$$s_{SFb}^{\varnothing} v = \lambda s_{SFt}^{\varnothing} v \qquad \dots (16)$$

ليكن $E = (e_1, e_2, \dots, e_N)^T$ و $N = (n_1, n_2, \dots n_j \dots, \dots, n_c)$ بحيث ان n_j يمثل عدد عينات التدريب المكانية في الصنف j . ستعرف مصفوفة الانشار بين الاصناف الضبابية ومصفوفة الاتشار الكلية الضبابية للعينات المركزية في فضاء الميزة $\hat{A}$ كالآتي:

$$s_{SFb}^{\varnothing} = \frac{1}{N}\sum_{i=1}^{c} n_i \left(\tilde{m}_i^{\varnothing} - \overline{m}^{\varnothing}\right)\left(\tilde{m}_i^{\varnothing} - \overline{m}^{\varnothing}\right)^T ||i-j||$$

$$= \frac{1}{N}\sum_{i=1}^{c} n_i \left(\tilde{\Phi}(x_j)U^T e_i\right)\left(\tilde{\Phi}(x_j)U^T e_i\right)^T ||i-j||$$

$$= \frac{1}{N}\tilde{\Phi}(x_j)U^T \sum_{i=1}^{c} n_i\, e_i e_i{}^T U\tilde{\Phi}(x_j)^T ||i-j||$$

$$= \frac{1}{N}(i-j)\tilde{\Phi}(x_j)U^T MU\tilde{\Phi}(x_j)^T (i-j)' \qquad \dots (17)$$

$$s_{Ft}^{\varnothing} = \frac{1}{N}\sum_{j=1}^{N}\left(\Phi(x_j) - \overline{m}^{\varnothing}\right)\left(\Phi(x_j) - \overline{m}^{\varnothing}\right)^T ||i-j||$$

$$= \frac{1}{N}\sum_{j=1}^{N}\left(\Phi(x_j)E_j\right)\left(\Phi(x_j)E_j\right)^T ||i-j||$$

$$= \frac{1}{N}\Phi(x_j)\sum_{j=1}^{N} E_j E_j{}^T \Phi(x_j)^T ||i-j||$$

$$= \frac{1}{N}\Phi(x_j)E\Phi(x_j)^T ||i-j||$$

$$= \frac{1}{N}(i-j)\Phi(x_j)\Phi(x_j)^T (i-j)' \qquad \dots (18)$$

وعلى اساس ذلك يمكن الحصول على المعادلة الآتية:

$$\max_a J(a) = (i-j)\frac{e^T \tilde{K} U^T MU\tilde{K} e}{e^T \tilde{K}\tilde{K} e}(i-j)' \qquad \dots(19)$$

11-معايير الدقة والتصنيف: [10] **(Accuracy and misclassification Creterion)**

الدقة هي درجة اقتراب القيمة المحسوبة او المقاسة من القيمة الفعلية ، ومن الضروري جداً حساب دقة هذه القياسات لضمان الحصول على نتائج دقيقة تقاس الدقة على مدى قرب القياسات المختلفة من بعضها البعض .

1- الدقة الكلية (Overall Accuracy)

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

هي متوسط المرجح لحساسية و دقة الاختبار اذ يتم حساب الحساسية المرجحة و الدقة المرجحة بواسطة مصفوفة الانتشارداخل الاصناف و بين الاصناف .

$$\text{Sensitivity} = \frac{a}{a+c} \quad , \quad \text{Specificity} = \frac{d}{b+d} \quad , \quad \text{Prevalence} = \frac{a+c}{N}$$

$$\text{Prevalence} = \frac{b+d}{N} \quad 1$$

جدول رقم (1) جدول التصنيف

	Positive	Negative
Positive	a	b
Negative	c	d
	a+c	b+d

N. يمثل عدد الحالات المصنفة بشكل صحيح من مجموع الحالات الكلية a+d

$$\text{Accuracy} = \frac{a+d}{N} = \left[\frac{a+c}{N}\right]\left[\frac{a}{a+c}\right] + \left[\frac{b+d}{N}\right]\left[\frac{d}{b+d}\right]$$

= **(**Prevalence) (Sensitivity) + (1– Prevalence) (Specificity)

الدقة الكلية (العامة) هي متوسط المرجح للحساسية والدقة باستعمال معادلة الدقة الكلية الواردة في الجدول 1 .

2- متوسط الدقة (Average Accuracy) [10]

متوسط الدقة (AA) هو متوسط دقة التصنيف لكل فئة.

$$AA = \frac{\sum_{i=1}^{m} CA_i}{m}$$

اذ ان :

C : تمثل مصفوفة الانتشار .

m : تمثل عدد الفئات

12- المحاكاة:

يمكن ايجاد مجموعة كبيرة من الاساليب لتقدير دالة الكثافة لمتغير عشوائي لامعلمي اذ أنتج الباحثون كمية هائلة من الأفكار التي تتوافق بشكل خاص مع مسألة اختيار معلمة عرض الحزمة في انموذج مقدر كثافة اللب لغرض تطبيقه في التحليل التمييزي اذ تناوله كثير من الباحثين في

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

سياق تقدير دالة كثافة اللب في تمهيد وتصنيف المشاهدات . اذ تم تكرارعدد التجربة لـ (1000) تكرار.

13- توليد مصفوفة الانتماء الضبابية (Fuzzy Membership Matrix)

تم توليد مصفوفة الانتماء الضبابية بموجب خوارزمية الجار الأقرب الصنف k الضبابية (Fuzzy k-Nearst Neighbor Algorithem) (FKNN) وباستعمال المعادلة رقم (11)

التجربة الاولى: h_v=700 و h_s=1

جدول (2) نتائج معايير الدقة والمقارنة بين اسلوب التحليل التمييزي اللبي المكاني واسلوب التحليل التمييزي الضبابي اللبي المكاني عندما تكون نسبة عينات التدريب 0.30 وعرض الحزمة h_v=700 و h_s=1

n	Method	SKDA	SKFDA
50	OA	71.15215	65.25547
	AA	70.34446	64.35659
Rank		1	2
200	OA	97.08535	97.88867
	AA	96.27766	96.99979
Rank		2	1
1000	OA	98.66305	99.46637
	AA	97.85536	98.57749
Rank		2	1

نلاحظ من جدول (2) عندما تكون نسبة عينات التدريب 0.30 وعرض الحزمة h_v=700 و h_s=1 مايأتي:

- **عند (n=50)** : حقق اسلوب التحليل التمييزي اللبي المكاني (SKDA) اعلى معايير للدقة (OA) و (AA) وجاء هذا الأسلوب في المرتبة الأولى , تلاه اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA)

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

- **عند (n=200, 1000)** :حقق اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) اعلى معايير للدقة (OA) و (AA) وجاء هذا الأسلوب في المرتبة الأولى , تلاه اسلوب التحليل التمييزي اللبي المكاني (SKDA) بالمرتبة الثانية .

جدول (3) نتائج معايير الدقة والمقارنة بين اسلوب التحليل التمييزي اللبي المكاني والأسلوب التحليل التمييزي الضبابي اللبي المكاني عندما تكون نسبة عينات التدريب 0.80 وعرض الحزمة $h_v=700$ و $h_s=1$

N	Method	SKDA	SKFDA
50	OA	67.25269	61.35601
	AA	66.44500	60.45713
Rank		1	2
200	OA	93.18589	93.98921
	AA	92.37820	93.10033
Rank		2	1
1000	OA	96.76359	97.56691
	AA	95.95590	97.67803
Rank		2	1

نلاحظ من جدول (3) عندما تكون نسبة عينات التدريب 0.80 وعرض الحزمة $h_v=700$ و $h_s=1$ مايأتي:

- **عند (n=50)** : حقق اسلوب التحليل التمييزي اللبي المكاني (SKDA) اعلى معايير للدقة (OA) و (AA) وجاء هذا الأسلوب المرتبة الأولى , تلاه اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) بالمرتبة الثانية .
- **عند (n=200, 1000)** :حقق اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) اعلى معايير للدقة (OA) و (AA) وجاء هذا الاسلوب المرتبة الأولى , تلاه اسلوب التحليل التمييزي اللبي المكاني (SKDA) بالمرتبة الثانية .

<u>التجربة الثانية</u> :$h_v=700$ و $h_s=3$

جدول (4) نتائج معايير الدقة والمقارنة بين اسلوب التحليل التمييزي اللبي المكاني والأسلوب التحليل التمييزي الضبابي اللبي المكاني عندما تكون نسبة عينات التدريب 0.30 وعرض الحزمة h_v=700 و h_s=3

n	Method	SKDA	SKFDA
50	OA	54.97215	51.08432
	AA	54.19442	50.08332
Rank		1	2
200	OA	64.98451	71.11089
	AA	65.20521	69.60311
Rank		2	1
1000	OA	80.14524	81.33232
	AA	80.32721	80.83132
Rank		2	1

نلاحظ من جدول (4) عندما تكون نسبة عينات التدريب 0.30 وعرض الحزمة h_v=700 و h_s=3 مايأتي:

- **عند (n=50)** : حقق اسلوب التحليل التمييزي اللبي المكاني (SKDA) اعلى معايير للدقة (OA) و (AA) وجاء هذا الأسلوب في المرتبة الأولى , تلاه اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) بالمرتبة الثانية .
- **عند (n= 200, 1000)** :حقق اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) اعلى معايير للدقة (OA) و (AA) وجاء هذا الاسلوب في المرتبة الأولى , تلاه اسلوب التحليل التمييزي اللبي المكاني (SKDA) بالمرتبة الثانية .

جدول (5) نتائج معايير الدقة والمقارنة بين اسلوب التحليل التمييزي اللبي المكاني والاسلوب التحليل التمييزي الضبابي اللبي المكاني عندما تكون نسبة عينات التدريب 0.80 وعرض الحزمة h_v=700 و h_s=3

n	Method	SKDA	SKFDA
50	OA	50.95853	47.07070
	AA	50.18080	46.06970
Rank		1	2
200	OA	60.97089	67.09727
	AA	61.19159	65.58949
Rank		2	1
1000	OA	76.13162	78.31870
	AA	76.31359	78.81770
Rank		2	1

نلاحظ من جدول (5) عندما تكون نسبة عينات التدريب 0.80 وعرض الحزمة h_v=700 و h_s=3 مايأتي:

- **عند (n=50)** : حقق اسلوب التحليل التمييزي اللبي المكاني (SKDA) اعلى معايير للدقة (OA) و (AA) وجاء هذا الأسلوب في المرتبة الأولى , تلاه اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) بالمرتبة الثانية .
- **عند (n= 200, 1000)** :حقق اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) اعلى معايير للدقة (OA) و (AA) وجاء هذا الاسلوب في المرتبة الأولى , تلاه اسلوب التحليل التمييزي اللبي المكاني (SKDA) بالمرتبة الثانية .

التجربة الثالثة: h_v=700 و h_s=6

جدول (6) نتائج معايير الدقة والمقارنة بين اسلوب التحليل التمييزي اللبي المكاني والاسلوب التحليل التمييزي الضبابي اللبي المكاني عندما تكون نسبة عينات التدريب 0.30 وعرض الحزمة h_v=700 و h_s=6

N	Method	SKDA	SKFDA
50	OA	67.12260	63.23477
	AA	66.34487	62.23377
Rank		1	2
200	OA	71.26546	68.36677
	AA	70.47687	69.36577
Rank		1	2
1000	OA	79.37734	77.37066
	AA	81.13956	76.59287
Rank		1	2

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

نلحظ من جدول (6) عندما تكون نسبة عينات التدريب 0.30 وعرض الحزمة h_v=700 و h_s=6 مايأتي:

- **عند (n=50)** : حقق اسلوب التحليل التمييزي اللبي المكاني (SKDA) اعلى معايير للدقة (OA) و (AA) وجاء هذا الأسلوب في المرتبة الأولى , تلاه اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) بالمرتبة الثانية .
- **عند (n= 200, 1000)** : حقق اسلوب التحليل التمييزي اللبي المكاني (SKDA)اعلى معايير للدقة (OA) و (AA) وجاء هذا الأسلوب في المرتبة الأولى , تلاه اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) بالمرتبة .

جدول (7) نتائج معايير الدقة والمقارنة بين اسلوب التحليل التمييزي اللبي المكاني والاسلوب التحليل التمييزي الضبابي اللبي المكاني عندما تكون نسبة عينات التدريب 0.80 وعرض الحزمة h_v=700 و h_s=6

n	Method	SKDA	SKFDA
50	OA	74.31781	70.42998
	AA	73.54008	69.42898
Rank		1	2
200	OA	77.56198	75.86067
	AA	79.56098	76.67208
Rank		1	2
1000	OA	83.57255	81.56587
	AA	82.33477	80.78808
Rank		1	2

نلاحظ من جدول (7) عندما تكون نسبة عينات التدريب 0.80 وعرض الحزمة h_v=700 و h_s=6 مايأتي:

- **عند (n=50)** : حقق اسلوب التحليل التمييزي اللبي المكاني (SKDA) اعلى معايير للدقة (OA) و (AA) وجاء هذا الأسلوب في المرتبة الأولى , تلاه اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) بالمرتبة .

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

- **عند (n= 200, 1000)** : حقق اسلوب التحليل التمييزي اللبي المكاني (SKDA) اعلى معايير للدقة (OA) و (AA) وجاء هذا الأسلوب في المرتبة الأولى , تلاه التحليل التمييزي الضبابي اللبي المكاني (SKFDA) بالمرتبة الثانية .

التجربة الرابعة: h_v=700 و h_s=10

جدول (8) نتائج معايير الدقة والمقارنة بين اسلوب التحليل التمييزي اللبي المكاني والاسلوب التحليل التمييزي الضبابي اللبي المكاني عندما تكون نسبة عينات التدريب 0.30 وعرض الحزمة h_v=700 و h_s=10

n	Method	SKDA	SKFDA
50	OA	65.33600	61.44817
	AA	64.55827	60.44717
Rank		1	2
200	OA	69.47886	66.58017
	AA	68.69027	67.57917
Rank		1	2
1000	OA	77.59074	75.58406
	AA	79.35296	74.80627
Rank		1	2

نلاحظ من جدول (8) عندما تكون نسبة عينات التدريب 0.30 وعرض الحزمة h_v=700 و h_s=10 مايأتي:

- **عند (n=50)** : حقق اسلوب التحليل التمييزي اللبي المكاني (SKDA) اعلى معايير للدقة (OA) و (AA) وجاء هذا الأسلوب في المرتبة الأولى , تلاه اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) بالمرتبة الثانية .
- **عند (n= 200, 1000)** : حقق اسلوب التحليل التمييزي اللبي المكاني (SKDA)اعلى معايير للدقة (OA) و (AA) وجاء هذا الأسلوب في المرتبة الأولى , تلاه اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) بالمرتبة الثانية .

جدول (9) نتائج معايير الدقة والمقارنة بين اسلوب التحليل التمييزي اللبي المكاني والاسلوب التحليل التمييزي الضبابي اللبي المكاني عندما تكون نسبة عينات التدريب 0.80 وعرض الحزمة h_v=700 و h_s=10

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

n	Method	SKDA	SKFDA
50	OA	63.31780	59.42997
	AA	62.54007	58.42897
Rank		1	2
200	OA	67.46066	64.56197
	AA	66.67207	65.56097
Rank		1	2
1000	OA	75.57254	73.56586
	AA	77.33476	72.78807
Rank		1	2

نلاحظ من جدول (9) عندما تكون نسبة عينات التدريب 0.80 وعرض الحزمة h_v=700 و h_s=10 مايأتي:

- **عند (n=50)** : حقق اسلوب التحليل التمييزي اللبي المكاني (SKDA) اعلى معايير للدقة (OA) و (AA) وجاء هذا الأسلوب المرتبة الأولى , تلاه اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) بالمرتبة الثانية .
- **عند (n= 200, 1000)** : حقق اسلوب التحليل التمييزي اللبي المكاني (SKDA)اعلى معايير للدقة (OA) و (AA) وجاء هذا الأسلوب المرتبة الأولى , تلاه اسلوب التحليل التمييزي الضبابي اللبي المكاني (SKFDA) بالمرتبة الثانية .

14-الإستنتاجات (Conclusions)

من خلال ما تم التوصل اليه من نتائج في الجانبين التجريبي والتطبيقي تم التوصل الى الاستنتاجات الآتية:

1) حققت اسلوب التحليل التمييزي الضبابي اللبي المكاني (**SKFDA**) افضلية على اسلوب التحليل التمييزي اللبي المكاني (**SKDA**) كلما زاد حجم العينة عند قيمة معلمة تمهيد مواقع المشاهدات h_s= 3.

2) عندما زادت قيمة معلمة تمهيد لمواقع المشاهدات h_s عن (3) قلت افضلية اسلوب التحليل التمييزي الضبابي اللبي المكاني (**SKFDA**) وزادت افضلية اسلوب التحليل التمييزي اللبي المكاني عند جميع الحجوم العينات.

3) حقق اسلوب التحليل التمييزي الضبابي اللبي المكاني (**SKFDA**) افضلية عند احجام العينات الكبيرة عند معلمة تمهيد مواقع المشاهدات h_s=1 ، اي عندما كانت قيمة معلمة التمهيد قليلة .

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (19) عدد (4) 2023

4) نلاحظ افضلية اسلوب التحليل التمييزي اللبي المكاني عند حجوم العينات الكبيرة و بقيم مناسبة لمعلمة تمهيد مواقع المشاهدات.

15-التوصيات (Recommendations)

من خلال الاستنتاجات التي تم التوصل اليها نوصي بالآتي:

1- عدم استعمال اسلوب التحليل التمييزي الضبابي اللبي المكاني في حالة كون العينات صغيرة .

2- استعمال دوال لبية اخرى غير الدالة الكاوسية مثل دالة (**Epankchove Kernel**) والمنتظمة (**Uniform Kernel**) في اسلوب التحليل التمييزي الضبابي اللبي المكاني لمعرفة سلوك التصنيف وفق هذه الدوال .

3- استعمال معلمة عرض الحزمة المناسبة واستعمالها مع اساليب اللبية الاخرى و مقارنتها مع اسلوب التحليل التمييزي الضبابي اللبي المكاني .

المصادر

1- Ahmed, M.S., N'diaye, M., Attouch, M.K., Dabo-Niang, S. 2018 . : k-nearest neighbors prediction and classification for spatial data. arXiv preprint arXiv :1806.00385

2- Andrienko, G. and Andrienko, N. (2001). Exploring spatial data with dominant attribute map and parallel co-ordinates. Computers, Environment and Urban Systems, 25, 5–15CrossRefGoogle Scholar.

3- Arbia, G., Griffith, D. and Haining, R. P. (2003). Spatial error propagation when computing linear combinations of spectral bands: the case of vegetation indices. Environmental and Ecological Statistics, 10 (to appear).

4- Aydilek, I.B. and Arslan, A. (2012). A novel hybrid approach to estimating missing values in databases using k-nearest neighbors and neural networks, International Journal of Innovative Computing, Information and Control 7(8): 4705–4717.

5- C. RENCHER, ALVIN, (2001), " Methods of Multivariate Analysis ", Second Edition, A JOHN WILEY & SONS, INC. PUBLICATION.

6- D . A . Griffith , Richardson . (2009) . " Spatial Autocorrelation" , university of Texas at Dallas , TX, USA

7- L., Li, J., Plaza, A., and Li, Y. (2017). Discriminative low-rank gabor filtering for spectral-spatial hyperspectral image classification. IEEE Trans. Geoscience and Remote Sensing, 55(3):1381–1395.

8- Liu, X., & Zhang, L. (2017). Kernel fuzzy discriminant analysis using a pairwise constraint for feature extraction. Pattern Recognition Letters, 90, 55-61.

9- Shekhar, S., Evans, M. R., Kang, J. M., and Mohan, P. (2011). Identifying patterns in spatial information: A survey of methods. Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery, 1(3):193–214.

10- Soumia Boumeddane1, Leila Hamdad1, Sophie Dabo-Niang2 and Hamid Haddadou1, (2019)," Spatial Kernel Discriminant Analysis: Applied for Hyperspectral Image Classification", DOI: 10.5220/0007372401840191 In Proceedings of the 11th International Conference n gents and Artificial Intelligence (ICAART 2019), pages 184-191 ISBN: 978-989-758-350-6 Copyrightc 2019 by SCITEPRESS – Science and Technology Publications, Lda

11- Thomas Ledl, (2004), " Kernel Density Estimation: Theory and Application in Discriminant Analysis ", AUSTRIAN JOURNAL OF STATISTICS Volume 33 (2004), Number 3, 267-279.

12- Vu, J., and L. Turner , (2008) Economic Significance of Tourism: A Discriminant Analysis. Tourism and Hospitality Research, Training and Practice, S. Richardson, L. Friedline, A. Patiar and M. Ternel, eds., pp.919-933. Gold Coast: Griffith University.